

eTranslation TermBank – naujas tarptautinis projektas Europos Komisijos automatinio vertimo kokybei pagerinti

EGLĖ PESLIAKAITĖ

Lietuvių kalbos institutas

2017 m. rugsėjo 1 d. pradėtas vykdyti naujas tarptautinis projektas „*eTranslation TermBank*: konkrečių sričių ir kalbų terminijos išteklių rinkimas ir pateikimas viešajam skaitmeninių paslaugų sektoriui“.

Pagrindinis projekto koordinatorius – *Tilde* (Latvija). Projekte taip pat dalyvauja aštuonių institucijų atstovai iš Lietuvos (Lietuvių kalbos institutas), Danijos (Kopenhagos universitetas), Estijos (Estų kalbos institutas), Slovėnijos (Jožefo Stefano institutas), Austrijos (Tarptautinis terminologijos tinklas (TermNet)), Islandijos (Arnio Magnussono islandų studijų institutas) ir Švedijos (Kalbos ir folkloro institutas, Švedijos terminologijos centras).

Projekto tikslas – pagerinti Europos Komisijos automatinio vertimo kokybę ir aprėptį.

AUTOMATINIS VERTIMAS – KALBOS BARJERAMS EUROPOJE MAŽINTI

Kalbų įvairovė – didelė Europos Sąjungos vertybė, tačiau šių dienų skaitmeniniame pasaulyje daugiakalbystė dažnai sukelia kliūčių, kalbos barjerai sunkina ne tik bendravimą, bet ir Europos Sąjungos piliečių ir šalių bendradarbiavimą.

Automatinis vertimas (angl. *eTranslation*) yra viena iš paslaugų, finansuojamų pagal Europos infrastruktūros tinklų priemonės (EITP) (angl. *Connecting Europe Facility (CEF)*) programą, siekiančią sujungti trūkstamas grandis tarp Europos energijos, transporto sričių ir skaitmeninių išteklių. Vertimo paslauga veikia nuo 2013 m. ir yra pasiekama ES institucijoms, valstybių narių administracijoms ir kelioms Europos Komisijos informacijos sistemoms bei interneto paslaugoms. Šis vertimas yra pagrįstas statistinio mašininio vertimo (MV) technologija, kuri mokosi versti iš žmonių atliktų turimų vertimų (plačiau žr. *European Language Resource Co-*

ordination. Europos kalbų išteklių koordinavimas, prieiga per internetą: <http://www.lr-coordination.eu>). Naudodamiesi automatiniu vertimu vartotojai iškart gali išversti žodžius, sakinius, dokumentus ir interneto svetaines iš vienos kalbos į kitą.

Svarbiausia, kad automatinio vertimo sistema taip pat gali būti prijungta prie bet kurios Europos bendradarbiavimui reikalingos viešojo administravimo skaitmeninių paslaugų infrastruktūros (angl. *Digital Service Infrastructures (DSIs)*), pavyzdžiui, elektroninių sveikatos paslaugų, atvirųjų duomenų portalo, elektroninių ginčų sprendimų portalo ir pan., ir padaryti šią infrastruktūrą daugiakalbę. Vis dėlto norint pasiekti efektyvių ir Europos viešųjų paslaugų administravimo institucijoms bei viešųjų paslaugų vartotojams, visuomenei naudingų rezultatų, t. y. kokybiško vertimo ir galimybės naudotis paslaugomis savo gimtąja kalba, reikia sukaupti kuo daugiau tinkamų kalbos išteklių. Siekdami šio tikslo ir susivienijo projekto *eTranslation TermBank* partneriai.

PROJEKTO UŽDAVINIAI IR PLANUOJAMI REZULTATAI

Projektą *eTranslation TermBank* prižiūri EITP remiamas Europos kalbų išteklių koordinavimo tinklas ir jai priklausantys Kalbų išteklių tarybos nariai (nacionaliniai atstovai kiekvienoje Europos Sąjungos šalyje), kurie teikia informaciją apie kiekvienos šalies terminijos išteklius.

Per pusantrų metų trukusią projektą numatyta rinkti konkrečių sričių terminijos išteklius visomis Europos Sąjungos (ir norvegų bei islandų) kalbomis ir juos pateikti viešajam skaitmeninių paslaugų sektoriui.

Šiuo metu Europos Komisija siekia sustiprinti kelių sričių teikiamas viešąsias skaitmenines paslaugas, kurioms reikalingas automatinis vertimas, t. y. sveikata (*e. sveikata*), verslo teisė (įmonių įregistravimas, išregistravimas (*e. teisė*)) ir vartotojų apsauga (*elektroninis ginčų sprendimas*).

Projektas bus įgyvendinamas 3 etapais:

1. Terminijos išteklių identifikavimas.
2. Identifikuotų išteklių įvertinimas.
3. Tinkamų terminijos išteklių teikimas automatinio vertimo platformai.

Kadangi svarbu, kad automatinis vertimas būtų naudingas visai Europai ir verstų iš visų Europos Sąjungos kalbų į visas, devyni projekto partneriai pasidalijo Europos kalbas, kurių terminijos ištekliai bus identifikuojami, kaupiami, įvertinami ir teikiami automatiniam vertimui. Lietuvos atstovai projekto metu dirbs su lietuvių, lenkų ir italų kalbų terminijos ištekliais.

Įgyvendindami projekto tikslus, partneriai bendradarbiaus su visos Europos nacionalinėmis institucijomis, dirbančiomis su terminija ir turinčiomis reikalingų išteklių.

Projekto metu numatoma surinkti 150 naujų, Europos Komisijos duomenų bazėse dar neužregistruotų ir į automatinio vertimo sistemą neįtrauktų terminijos išteklių visomis Europos Sąjungos bei norvegų ir islandų kalbomis. Projekto koordinatorių iniciatyva automatinio vertimo sistema bus papildyta naujaisiais terminijos ištekliais, ir taip bus pagerinta automatinio vertimo kokybė bei išplėsta galimybė visavertiškai naudotis Europos Komisijos teikiamomis viešosiomis paslaugomis. Taip pat bus sudaryti nacionalinių institucijų, dirbančių su terminija, ir konkrečias sritis apimančių terminijos išteklių katalogai, pateikta terminijos išteklių automatiniam vertimui rengimo ir taikymo metodologija. Taigi projektas taip pat bus itin vertingas siekiant plėsti ir stiprinti Europos terminologų bei institucijų, dirbančių su terminijos ištekliais, bendradarbiavimą.

eTranslation Termbank – A New International Project to Improve the Quality of Automated Translations for the European Commission

1 September 2017 saw the start of a new international project called *eTranslation Termbank: collection and provision of terminological resources in sector-specific domains and languages of interest to sector-specific digital public services*.

The key coordinator for this project is *Tilde* (Latvia); plus, there are eight other partners involved in the project: the Institute of the Lithuanian Language (Lithuania), the University of Copenhagen (Denmark), the Institute of the Estonian Language (Estonia), the Jožef Stefan Institute (Slovenia), the International Network of Terminology, TermNet (Austria), the Árni Magnússon Institute for Icelandic Studies (Iceland), and the Institute of Language and Folklore and the Swedish Centre for Terminology (Sweden).

The project aims to improve the quality and coverage of automated translations for the European Commission.

ETRANSLATION TO BRING DOWN LANGUAGE BARRIERS IN EUROPE

Linguistic diversity is a key value of the European Union, however in today's digital world multilingualism often creates obstacles and language barriers that make it more difficult for the citizens and Member States of the European Union to communicate and cooperate.

Automated translation, or eTranslation, is a service financed under the Connecting Europe Facility (CEF) that aims to provide the missing links between energy, transport domains, and digital resources in Europe. The translation service was launched in 2013 and is available to EU institutions, Member State administrations, and several European Commission information systems and Internet services. This type of translation is based on the machine translation (MT) technology, which learns to translate from the available pool of man-made translations (for details, see European Language Resource Coordination, online access at: <http://www.lr-coordination.eu>). Automated translation allows users to translate words, whole sentences, documents, and websites from one language to another.

The most important thing is that the eTranslation system can also be customised and connected to render multilingual any public administration digital service infrastructure (DSI) needed to enable cooperation in Europe, such as e-health services, the open data portal, the Online Dispute Resolution (ODR) portal, and so on. Still, to achieve outcomes that are both effective and beneficial to European public service administration institutions and public service users, which outcomes would constitute high-quality translations and a possibility to use the services in one's mother tongue, we have to accumulate as many adequate language resources as possible. It is to attain that purpose that the partners of the *eTranslation TermBank* project joined forces.

PROJECT OBJECTIVES AND DELIVERABLES

The *eTranslation TermBank* project is supervised by the European Language Resource Coordination network and its members from the Language Resources Council (national representatives in every Member State) supporting the collection of data in each country.

The eighteen-month project is geared towards collection and provision of sector-specific terminological resources for all European Union (as well as Norwegian and Icelandic) languages to sector-specific digital public services.

Currently, the European Commission seeks to strengthen the digital public services in several domains that require automated translation, such as health

(e-Health), business legislation (company registration and de-registration; e-Justice), and consumer protection (ODR - Online Dispute Resolution)).

The project will be implemented in three stages:

1. Identification of the required terminological resources.
2. Assessment of the resources identified.
3. Provision of the appropriate terminological resources to the eTranslation platform.

Since it is critical for automated translation to benefit the entire Europe and be able to provide translations from and to every European Union language, the nine partners in the project have divided among themselves the European languages whose resources they will be identifying, compiling, and providing for automatic translation. The Lithuanian party will work with Lithuanian, Polish, and Italian language terminological resources in this project. To achieve the objectives of the project, the partners will cooperate with pan-European national bodies that work with terminology and possess the required resources.

A total of 150 new terminological resources for all European Union languages, as well as Norwegian and Icelandic, that have never been registered in the European Commission's databases or included in the automated translation system are to be collected in the course of implementation of the project. At the initiative of the project coordinators, the eTranslation system will be enhanced with the new terminological resources, which will allow improving the quality of automated translation and secure the opportunity for the users to fully enjoy the European Commission's public services. In the process of delivering the project's objectives, a catalogue of national institutions working with terminology and a catalogue of sector-specific terminological resources will be made and a methodology of preparation and application of terminological resources for automated translation drafted, which will greatly increase the project's value for expanding and reinforcing the fine-tuned cooperation among European terminologists and institutions that work with terminological resources.

Straipsnį į anglų kalbą vertė Jūras Barauskas

Gauta 2017-10-30

Eglė Pesliakaitė
Lietuvių kalbos institutas
P. Vileišio g. 5, LT-10308 Vilnius
El. p. egle.pesliakaite@lki.lt